Available online at www.sciencedirect.com

SciVerse ScienceDirect

Energy Procedia 16 (2012) 1913 – 1918

Energy Procedia

2012 International Conference on Future Energy, Environment, and Materials

Principal Component Analysis of Electricity Consumption Factors in China

Jing Zhang, Xin-yao Yang, Fei Shen, Yuan-wei Li, Hong Xiao, Hui Qi, Hong Peng, Shi-huai Deng*

Provincial Key Laboratory of Agricultural Environmental Engineering,
Sichuan Agricultural University,
Ya'an, Sichuan, P.R. China, 625014
**Corresponding author*

Abstract

This study aims to identify the principal components in ten prevailing factors, which would account for the influence of major economic variables on electricity consumption. Principal component analysis on these variables suggests two principal components, which respectively have Eigen-values of 8.28 and 1.04 and a cumulative variance of 82.77% and 93.19%. The regression results between electricity consumption and two principal components indicate close match between the calculated and the actual electricity consumption. The first principal component positively impact on electricity consumption, while the second one has negative influence. According to the coefficient of each variable in the combination, we found that gross domestic product, income, industrial output value, commodity exports, added services industry value, household numbers, percentage of rural population and price index impact positively on electricity consumption, while increase in commodity imports and efficiency of electricity reduces electricity consumption. These imply the possibility to control the fast growth of electricity consumption by imposing policies to revise the concerned factors.

Keywords: principal component analysis; regression analysis; electricity consumption; economic development;

1. Introduction

Energy saving is a key pivot for sustainable development. Electricity as a major energy type plays an unprecedented role in social development. Improvement in electricity saving requires identification and control of the factors that have dominant impact on electricity consumption. A number of developed electricity forecasting models describe the impact of economy[1-3], population[4-7] and technology[8-9]

 doi:10.1016/j.egypro.2012.01.292

factors on electricity consumption using regression analysis techniques. While calculation results employing these empirical or regression-based models often displayed good correlation between electricity consumption and the primary variables, a number of them either fail to consider certain important variables, or do not adequately eliminate bias error in the economic variables incurred during the multiple linear regression analysis[10-11]. The primary goal of the present work is to identify the main factors influencing the electricity consumption in China.

2. Principal components analysis (PCA)

In analyzing long-term meteorological variables, it is often advantageous to group the key variables directly affecting electricity consumption. PCA is a multivariate statistical technique for quantitatively describing the degree of inter-dependency for a set of correlated variables[12-13]. This technique excels in elucidating the casual relationship for its capacity to group the complex and highly correlated meteorological variables. PCA conducted on centered data or anomalies allow the pattern of simultaneous variations to be identified. This permits to reduce a data set consisted of a large number of inter-correlated variables to that of fewer hypothetical and uncorrelated components, which nevertheless represent the major fraction of the variability in the original data. These components result from linear combinations of the original variables, with coefficients given by the eigenvector, and each component contributes to the total variance of the original variables. The analysis scheme requires that the components be ranked in descending order according to the magnitude of their contributions such that the first component, responsible for the biggest variance, explains the greatest variance of the original variables, the second for the next largest, and so on.

Based on literature, we selected 10 variables (Data on electricity consumption, gross domestic product, income, industrial output value, commodity exports, commodity imports, and added industry services value are from International Energy Agency (IEA). Data on household number, percentage of rural population, and efficiency of electricity are from National Bureau of Statistics of China (NBSC)), namely, gross domestic product (x_1), income (x_2), industrial output value (x_3), commodity imports(x_4), commodity exports (x_5), added services industry value (x_6), household numbers (x_7), percentage of rural population (x_8), efficiency of electricity (x_9) and price index(x_{10}). Measurement data for twenty two years (1985-2007) were gathered for this study. Table 1 summarizes the results of correlation between all variables. About 53.33% and 80.00% of correlation coefficients among the variables are over 0.9 and 0.6, respectively.

Table 2 summarizes the ten principal components for PCA. Each Eigen-value measures the variance accounted for by the corresponding principal component, and calculation of all possible Eigen-values permits all the variance of the original variables to be quantified. Principal components can be ranked according to their ability to explain variance in the original data set. A common approach is to select only those with Eigen-values equal to or greater than one or with at least 80% cumulative variance[14].

Based on the above principle, the first and second principal components were retained since they had Eigen-values greater than one and a cumulative variance above 80% (table 2). Table 3 lists two new sets of variables, Z_1 and Z_2, which are calculated as a linear combination of the original ten variables with their respective coefficients.

$$Z_1 = \sum \alpha_{1i} x_i \tag{1}$$

$$Z_2 = \sum \alpha_{2i} x_i \tag{2}$$

Where Z_1 and Z_2 are the first and second principal components of the ten variables. α_{1i} and α_{2i} are the coefficients of each variable for its corresponding principal component.

Table 1. The results of variables' correlations

Variables	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
x_1	1.00	0.95	1.00	-0.94	0.99	0.99	0.69	1.00	-0.36	0.92
x_2	0.95	1.00	0.95	-0.92	0.89	0.89	0.68	0.94	-0.35	0.91
x_3	1.00	0.95	1.00	-0.94	0.99	0.99	0.69	1.00	-0.35	0.92
x_4	-0.94	-0.92	-0.94	1.00	-0.89	-0.91	-0.75	-0.94	0.36	-0.97
x_5	0.99	0.89	0.99	-0.89	1.00	1.00	0.66	0.99	-0.31	0.87
x_6	0.99	0.89	0.99	-0.91	1.00	1.00	0.69	0.99	-0.31	0.88
x_7	0.69	0.68	0.69	-0.75	0.66	0.69	1.00	0.70	0.03	0.75
x_8	1.00	0.94	1.00	-0.94	0.99	0.99	0.70	1.00	-0.36	0.91
x_9	-0.36	-0.35	-0.35	0.36	-0.31	-0.31	0.03	-0.36	1.00	-0.23
x_{10}	0.92	0.91	0.92	-0.97	0.87	0.88	0.75	0.91	-0.23	1.00

Table 2. Summary of principal component analysis

Component	Initial Eigen-values		
	Total	Variance (%)*	Cumulative (%)
1st	8.28	82.77	82.77
2nd	1.04	10.42	93.19
3rd	0.37	3.71	96.90
4th	0.19	1.92	98.82
5th	0.10	1.01	99.83
6th	0.01	0.13	99.96
7th	0.00	0.03	99.99
8th	0.00	0.01	100.00
9th	0.00	0.00	100.00
10th	0.00	0.00	100

*The percentage is given by the ratio of the individual Eigen-value to the trace of the correlation matrix.

3. Regression Analysis and Results

In the regression analysis we selected the electricity consumption (Y) as a dependent variable, and Z_1 and Z_2 as independent variables. We standardized the original variable using equation (3).

$$x' = (x - x'')/\sigma^2 \tag{3}$$

Where x' is the standardized data, x is original data including Y, Z_1, and Z_2, x'' is average of original data and σ is the standard deviation of original data.

Ordinary Least Square (OLS) regression method was employed to estimates the parameters in equation (4).

$$Y' = \mathrm{a}\, Z'_1 + \mathrm{b}\, Z'_2 \tag{4}$$

Where Y', Z'_1 and Z'_2 are the standardized data of Y, Z_1, and Z_2, respectively; a and b are the coefficients of the first and second principal components, respectively.

We evaluated the performance of the regression models. An error analysis was conducted by comparing the calculation results determined from equation (4) with the actual electricity consumption. In order to quantify the deviation, mean bias error (*MBE*) and root mean square error (*RMBE*) were calculated using equation (5) and equation (6).

$$MBE = (\sum (R_j - Y_j))/n \tag{5}$$

$$RMBE = ((\sum (R_j - Y_j)^2)/n)^{1/2} \tag{6}$$

Where R_j is the electricity consumption in year *j*, which is calculated using equation (4); Y_j is the actual electricity consumption in year *j* and n is the number of year, which is 22.

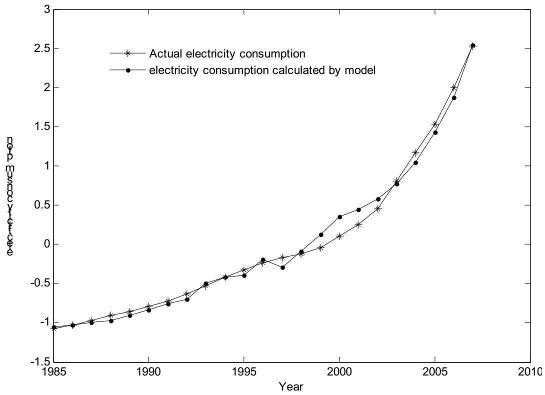

Fig.1 the calculated and actual electricity consumption

Table 3 and Fig.1 show the results of regression analysis. Parameters a and b pass the t-test at a significance level of 0.005. The F-test value of equation 4 is higher than the critical value, where the fiducial probability is blow 95%. The MBE value is around zero, suggesting a close match of the calculated with the actual electricity consumption. A RMBE value of 0.101 further confirms the high accuracy in calculation. Fig.1 compares the curve of actual consumption with that of the estimated consumption. Except for year 1998-2002, electricity consumption calculated by equation (4) is comparable to the actual value.

Table 3. Result of OLS Regression

Parameters	a	t-test	b	t-test
Value	0.982	43.3	-0.156	6.88
Parameters	R^2	F-test	MBE	RMBE
Value	0.990	2017	1.14×10^{-5}	0.101

4. Discussion

Regression analysis results show that when the first principal component increases by one unit, the dependent variable (electricity consumption) rises by 0.982 units. Conversely, when the second principal component increases by one unit, electricity consumption declines by 0.156 units. In order to discuss the influence of individual variables on electricity consumption, equation (7), derived from equation (1), equation (2) and equation (4), was used for further analysis.

$$Y' = \sum (a\alpha_{1i} + b\alpha_{2i})\, x_i \tag{7}$$

Table 4. Coefficient of variables

Variables	1st Component	2nd Component	$a\alpha_{1i}+b\alpha_{2i}$
x_1	0.111	-0.058	0.118
x_2	0.107	-0.052	0.113
x_3	0.111	-0.055	0.118
x_4	-0.115	0.021	-0.116
x_5	0.112	-0.035	0.116
x_6	0.115	-0.025	0.117
x_7	0.174	0.413	0.106
x_8	0.112	-0.054	0.118
x_9	0.126	0.857	-0.010
x_{10}	0.135	0.094	0.118

Table 4 lists the coefficient ($a\alpha_{1i}+b\alpha_{2i}$) of each variable. An increase, by one unit, of x_1, x_2, x_3, x_5, x_6, x_7, x_8 and x_{10} would result in a rise in electricity consumption by 0.118, 0.113, 0.118, 0.116, 0.117, 0.106, 0.118 and 118 units, respectively. In other words, these eight factors exert a positive impact on electricity consumption. On the other hand, a rise of x_4 and x_9 by one unit brings a decline in electricity consumption by 0.116 and 0.010 units, respectively, suggesting that commodity imports and efficiency of electricity are two factors having negative impact on electricity consumption.

Of the ten variables, except for the percentage of rural population, the other variables should grow continuously as a result of economic growth and social development. Therefore, it is reasonable to slow down the rising rate of electricity consumption by reducing the percentage of rural population, increasing the commodity exports, and increasing the efficiency of electricity.

5. Conclusions

(1) Principal component analysis of ten prevailing factors affecting the electricity consumption was conducted. The first and second principal components had Eigen-values greater than one and a cumulative variance of 82.77% and 93.19%, respectively.

(2) Regression model was used to calculate electricity consumption which was a dependent variable while the first and second principal components were selected as independent variables. The coefficient of determination (R^2) is 0.982 for the calculated and actual electricity consumption. Error analysis suggested close match of the calculated with the actual electricity consumption.

(3) According to the coefficient of each variable in the combination, the growth of gross domestic product, income, industrial output value, commodity exports, added services industry value, household numbers, percentage of rural population and price index would cause electricity consumption to increase. Conversely, increase in commodity imports and efficiency of electricity would make electricity consumption to decline.

References

[1]Z. Mohamed, P. Bodger, "Forecasting electricity consumption in New Zealand using economic and demographic variables", Energy, Jun. 2004, vol. 30, pp. 1833-1843, doi:10.1016/j.energy.2004.08.012.

[2]F. Egelioglu, AA. Mohamad, "Economic variables and electricity consumtion in Northern Cyprus", Energy, vol. 26, Apr. 2001, pp. 355-362, doi:10.1016/S0360-5442(01)00008-1.

[3]JL. Harris, L. Liu, "Dynamic structural analysis and forecasting of residential electricity consumption", International Journal of Forecasting, vol. 9, Dec. 2001, pp. 437-455, doi:10.1016/0169-2070(93)90072-U.

[4]M. Rajan, VK. Jain, "Modeling of electricity energy consumption in Delhi", Energy, vol. 24, Apr. 1999, pp. 351-361, doi:10.1016/S0360-5442(98)00087-5.

[5]XQ. Liu, BW. Ang, TN. Goh, "Forecasting of electricity consumption: a comparision between an economic model and a neural network model", In: IEEE international Conference on Neural Networks, vol. 2, Jun. 1991, pp. 1254-1259, doi: 10.1109/IJCNN.1991.170569.

[6]AZ. Al-Garni, SM. Zubair, JS. Nizami, "A regression model for electric-energy-consumption forecasting in Eastern Saudi Arabia", Energy, vol. 19, Oct. 1994, pp. 1043-1049, doi:10.1016/0360-5442(94)90092-2.

[7]S. Kadoshin, T. Nishiyama, I. Toishihide, "The trend in current and near future energy consumtion consumption from a statistical perspective", Application Energy, vol. 67, Dec. 2000, pp. 407-417, doi:10.1016/S0306-2619(00)00033-7.

[8]U. Able-Thomas, "Models of renewable energy technology transfor to developing countries", Renew Energy, vol. 9, Sep. 1996, pp. 1104-1107, doi:10.1016/0960-1481(96)88471-0.

[9]VB. Amatya, M. Chandrashekar, JB. Robinson, "Residential sector energy-supply-demand analysis: a modeling approach for developing countries and fuelwood-supply sustainablilty in Nepal" Energy, vol. 18, Apr. 1993, pp. 341-354, doi:10.1016/0360-5442(93)90069-P.

[10]K. Rijal, NK. Bansal, PD. Grover, "Rural household energy demand modeling, A case study of Nepal", Energy Economics, vol. 4, Oct. 1990, pp. 279-288, doi:10.1016/0140-9883(90)90018-B.

[11]RE. Abdel-Aal, AZ. Al-Farni, YN. Al-Nassar, "Modeling and forecasting monthly electric energy consumption in Eastern Saudi Arabia using abductive network", Energy, vol. 22, Sep. 1997, pp. 911-921, doi:10.1016/S0360-5442(97)00019-4.

[12]DS. Wilks, Statictical method in the atmospheric sciences: an introduction. San Diego: Academic Press, 1995.

[13]HV. Storch, FW.Zwiers, Statistical analysis in climate research. Cambridge: Cambridge University Press, 1999.

[14]JC. Lam, KK.W. Wan, S.L. Wong, TN.T. Lam, "Principal component analysis and long-turn building energy simulation correlation", Energy Conversion and Management, vol. 51, Jan. 2010, pp. 135-139, doi:10.1016/j.enconman.2009.09.004.